

# Practice-Based Evidence: Benchmarking NHS Primary Care Counselling Services at National and Local Levels

Chris Evans,<sup>1\*</sup> Janice Connell,<sup>2</sup> Michael Barkham,<sup>2</sup>  
Chris Marshall<sup>2</sup> and John Mellor-Clark<sup>2</sup>

<sup>1</sup>Rampton Hospital, Nottinghamshire Healthcare NHS Trust and Tavistock & Portman NHS Trust, UK

<sup>2</sup>Psychological Therapies Research Centre, University of Leeds, Leeds, UK

There are a number of problems for evidence-based practice (EBP) including limited generalizability of efficacy research results, the consequent lack of confidence in the relevance of such research, and the conceptual distance of most practitioners from the research process. The result is that EBP, although sound in principle, often fails to achieve its aim of improving practice. Practice-based evidence (PBE) provides a complementary bridge for the gap between research and practice to offset some of these problems, promoting collaboration between mental health services and academic institutions. This paper presents the initial results of such a collaboration via three phases: (1) the development of a referential database for primary care counselling services, (2) 'practitioner-friendly' feedback on grouped data to services, and (3) the combination of the two to build an evidence base for work with ethnic minorities—an area in which research trials are not well adapted to provide much evidence. Copyright © 2003 John Wiley & Sons, Ltd.

## INTRODUCTION

The evidence-based practice (EBP) paradigm has become increasingly influential in all health care. In this paradigm, primacy is generally given to evidence derived from systematic reviews, meta-

analytic reviews and randomized controlled trials (RCTs). In the UK, the output of this paradigm for psychological therapies in primary care has been considerable including high quality RCTs (e.g. Ward *et al.*, 2000), systematic reviews (e.g. Bower, Richards, & Lovell, 2001; Rowland *et al.*, 2000) and the development of practice guidelines for use in primary care (e.g. Department of Health, 2001). These products are central components in the overall plan for an appropriate knowledge base for the psychological therapies as set out in the strategic review of psychotherapy (Department of Health, 1996).

In principle the evidence-based practice paradigm should empower practitioners to answer locally pertinent questions, clinical or managerial, through critical appraisal of the best available

\*Correspondence to: Dr C. Evans, Rampton Hospital, Nottinghamshire Healthcare NHS Trust, Retford, Notts. DN22 0PD, UK.

E-mail: chris@psyctc.org

Contract/grant sponsor: Artemis Trust.

Contract/grant sponsor: Mental Health Foundation.

Contract/grant sponsor: Leeds Community and Mental Health Trust.

Contract/grant sponsor: Nottinghamshire Healthcare NHS Trust.

Contract/grant sponsor: Tavistock & Portman NHS Trust.

evidence. However, many of the constituent components of this paradigm have been criticized, including the philosophical and political assumptions underpinning EBP (e.g. Colyer & Kamath, 1999), the randomized controlled trial (e.g. Marshall, 2002), the limitations to the paradigm (e.g. Trinder, 2000) and user concerns (e.g. Faulkner & Thomas, 2002).

One specific constraint is that strict study inclusion criteria restrict the data available for analysis and thus limit the conclusions which can be drawn and constrain the generalizability of findings. While there is considerable evidence supporting the efficacy of the psychological therapies (e.g. Lambert & Ogles, 2003; Roth and Fonagy, 1996), there is still relatively little evidence for specific age groups, ethnic minorities, or large service areas, such as primary care. Data sufficient for precise estimation of effects for small subpopulations requires the availability of much larger datasets than those which will accumulate through prioritized RCTs and traditional research-driven studies.

Another more general problem for the 'pure' EBP paradigm is that practitioners need skills and confidence to use the model. In the psychological therapies, the considerable real challenges to the generalizability of RCT evidence give many practitioners serious doubts about, and poor confidence in, the applicability of EBP evidence to their own practice. A more prosaic problem is that unfamiliarity with quantitative data *per se* and with the measures reported, distances some practitioners from the paradigm. Critical appraisal and EBP trainings, while vital, can give practitioners skills but no data to address their local issues.

A complementary paradigm to EBP which addresses many of these concerns is that of practice-based evidence (PBE: Barkham & Mellor-Clark, 2000; Margison *et al.*, 2000). This paradigm uses an evidence-base derived from routine practice settings rather than from efficacy studies. The PBE paradigm should engage practitioners in the collection and ownership of data and in analyses of that data which can inform their practice. Two key components are central to the practice-based paradigm: effectiveness and practice (Barkham & Mellor-Clark, this issue; NAMHC, 1999). The effectiveness component addresses generalizability of results *across* particular services and settings. It does not provide strong causal attribution but it addresses generalizability and enables location of the activities and outcomes of a particular service within the range of data from other services. The practice component addresses the agenda of

analysing results *within* a service or setting. That is, it gives the ability to drill down into the data to ascertain individual differences and variations in relation to client subgroups. The present study provides a case study of practice-based investigation for one counselling service indicating how local data can be used within a service.

This model extends the traditional model of practice-based evidence in the psychological therapies which has been one of narrative reports with increasing self-reflection, context or cross-linkage to other data (*intra alia*: Eynon, 2001; Freud, 1977; Kohut 1979; Parry, Shapiro, & Firth 1986; Tustin 1958). This has great strengths, particularly when coupled with observational theories, recording systems or formal qualitative analytic methods. However, one clear weakness of such methods has been the lack of generalizability. Another traditional approach has been for individual services to collect in-house quantitative data (e.g. Bentovim, Boston, & Van El, 1987; Daryanani, Hindley, Evans, Fahy, & Turk, 2001; Ford *et al.*, 1990; Kramers, Evans, Dolan, Hume, & Lacey, 1991). However, such data are often collected more for local audit and managerial purposes and not shared in the research literature. Still another approach has involved regular case-by-case or service-wide use of self-report or rated measures (e.g. Brooker *et al.*, 1994; Luborsky, McLeuan, Woody, & O'Brien, 1985; Ryle, 1980; Treasure *et al.*, 1994). These latter approaches, of audit and use of outcome measures, can be of limited value in assessing a single service. However, they gain enormously if databases from individual services adopt common approaches to data collection and more so if such datasets are then collated to provide large or national referential datasets (Barkham *et al.*, 1998). This paper takes forward this argument by presenting multi-site referential data against which many or most services could compare their own data.

Collecting referential data and enabling it to be used meaningfully by local services requires collaborative action and agreement across two phases: issues of technology (i.e. measurement tools); and procedures for utilizing the data (i.e. feedback of information). For the former, identifying a common measurement tools is crucial. This need drove the development of the measures in the CORE system to provide a common minimum set of measures. The second need is also addressed by the CORE system within which we are developing coherent systems for reporting on the pre- and post-therapy context and change (Barkham *et al.*,

1998; Evans *et al.*, 2000; Mellor-Clark, Barkham, Connell, & Evans, 1999). The CORE system was used by the particular service and many other services who all contributed to the referential dataset reported on in this paper.

To progress beyond the mere collection of data requires feedback to services of useful parameters derived from the collected data. This is a relatively new area of activity and a recent systematic review of routine outcome measures showed no clear benefit from their use (Gilbody, House, & Sheldon, 2001). However, randomized trials of outcomes feedback to patients and practitioners are being carried out (e.g. McCabe & Priebe, 2002) and controlled studies of feedback to patients in psychotherapy have shown promise although revealing considerable variance (e.g. Lambert *et al.*, 2001). One key focus in the CORE battery project has been to provide reports to services which might help contribute to service management.

In this paper, we describe how PBE data can yield useful information both at the national and the service level. We recommend combining such analyses with careful inspection of data at the individual client level using the methods of reliable and clinically significant change (Evans, Margison, & Barkham, 1998; Jacobson & Truax, 1991), but this paper concentrates on grouped data. The paper is divided into three sections. First, we briefly describe the CORE system and the subsequent development of a national practice-based dataset. Secondly, through presentation of results for one particular service and the comparison of their data with the national referential data from other similar services, we illustrate how feedback to the service can be used to inform service policy and improve service delivery. Thirdly, as this service provides for an ethnically diverse population, ethnicity was important and is explored in some detail. Use of the service and referential datasets enabled us to explore issues which were of particular salience to that particular service and for which little evidence was available in the literature.

## THE PRACTICE-BASED NATIONAL REFERENTIAL DATASET

The aim of this section is to establish referential data for a range of service parameters that would provide context for individual services' data. This provides for the investigation of common and specific effects in counselling practice and delivery.

## Method

A database was created through the accumulation of data from a data mounting, analysis and reporting service based at the Psychological Therapies Research Centre (PTRC), University of Leeds.

## Measures

We used the two released measures within the CORE System battery: the CORE-OM and CORE-A.

*CORE-OM (Clinical Outcomes in Routine Evaluation-Outcome Measure; Evans et al., 2000, 2002).* This is a 34-item self-report measure primarily completed at pre- and post-therapy. Items cover the domains of subjective well-being, problems, and life/social functioning. In addition, it contains four items on risk to self, and two items on risk to others. Good internal reliability, sensitivity to change, test-retest stability, convergent validity in relation to other measures and discrimination between clinical and non-clinical populations have all been reported (Evans *et al.*, 2002).

*CORE-A (Clinical Outcomes in Routine Evaluation-Assessment; Mellor-Clark et al., 1999).* This comprises two practitioner-completed forms: the Therapist Assessment Form (TAF) and the End of Therapy Form (EOT). These forms were developed as an audit and evaluation tool that could be used alone or, ideally, as a complement to the CORE-OM. Data includes: intake demographics (age, gender, ethnic origin, employment, relationships/support); waiting times (dates of referral, assessment, and first appointment); intake severity and case mix (presenting problems and risk); medication; outcome of assessment (e.g. accepted, referred on); therapy process (type, mode, number of sessions and length of therapy); therapy ending (planned/unplanned); contextual factors (motivation, psychological mindedness, working alliance) and effectiveness (comparison of severity levels of presenting problems and risk at assessment and end of therapy; changes to medication, benefits of therapy).

## Procedure

The CORE measures are copyleft, i.e. they may be photocopied freely provided that they are not changed in any way. Partly because of this low cost, they are now widely used in Britain and increasingly in North America without any specific support for analyses. However, the CORE system development was always driven by a recognition

that most services would need help with analyses of their data. This paper reports on the first supported analysis that was designed as a crucial complement to the measures in the full CORE system in which the measures, which had been designed to be computer scanned, were batch processed for services. Another method of supported analysis is now also available using a licensed PC software package (CORE-PC) drawing on data accumulated through the batch process to provide referential parameters. The two support methods will appeal differently to different services.<sup>1</sup>

In the batch processing system, which underpins this paper, services sent completed batches of CORE system forms, covering a minimum period of 6 months, for analysis and reporting, which included a comparison of their service data with that of similar services. Data mounting was automated by the Formic™ system which exports the data in SPSS data files which are then checked thoroughly for scanning and data entry errors. Services agreed to the accumulation of anonymized data into a cumulative database. Each service was given a 3-h training session which included an introduction to the system, its rationale and advice on completion of the forms. In addition, each participating practitioner was provided with a comprehensive user manual that contained scoring information and guidelines for completion of the CORE system measures (CORE System Group, 1999). Services also had telephone support from the CORE team to deal with specific queries if required.

### Analyses

The data reported to services were expressed as simple means or proportions and in this paper we provide 95% confidence intervals (see e.g. Gardner & Altman (1986) and Rothman (1978) for the simple calculation of confidence intervals for a proportion from the observed proportion and  $n$ ; and for means from the observed mean, SD and  $n$ ). CORE-OM analyses are reported for the overall item mean (probably the most widely used scoring) and for the 28 'non-risk' as these items form a clean psychometric cross-sectional scaling whereas the risk items, as intended, show a clearly somewhat different dimensionality (see Evans *et al.*, 2002). Pro-rating of up to 10% of items

was used, i.e. up to three missing items were replaced with the mean of the rest for both overall and non-risk scores.

## Results

### Primary Care Referential Data

The main dataset as of end of 2002 comprised data for over 12000 clients from a wide variety of psychotherapy services. From this dataset, cases were selected which came from NHS primary care counselling services only, i.e. all services offering non-NHS primary or any secondary or tertiary services were excluded. This referential dataset contained data for 6610 clients from 33 NHS primary care services. Data from all 33 sites were aggregated and anonymized so no individual site is identifiable. The contribution of data from the different services ranged widely, from 40 clients from a relatively small service to 932 from one Trust. The mean number of clients per service was 200 (SD = 204; median = 135). Women accounted for 71.5% of the sample and the age range was from 11 to 89 years with a mean of 38 years (median = 37 years; SD = 13.1 years). The referral source was recorded for 5622 referrals (85.1%) of which 93.8% were from the GP or primary care practice.

Table 1 presents the major point parameters from the referential data together with 95% confidence intervals. Demographic data showed 4.5% of people attending primary care counselling services to be 20 years old or younger and 6.5% 60 years or older; 9% to be from ethnic minority groups and just over one-quarter to be living alone. Problem mix data showed about three-quarters to have experienced some level of depression and about the same for anxiety and about 60% to have experienced interpersonal difficulties. One in five clients were reported as having some level of risk to self and the same for suicide risk and about the same proportion presented with severity levels *below* an established clinical cut-off level on the self-report measure.<sup>2</sup> The mean waiting time to first

<sup>1</sup> Enquiries about CORE-PC should be directed to Core Information Management Systems Ltd, 47 Windsor Street, Rugby, CV21 3NZ; <http://www.coreims.co.uk/>; Tel: 01788 546019. E-mail: [j.mellor-clark@freeuk.com](mailto:j.mellor-clark@freeuk.com) Enquiries about batch processing should be directed to JC and MB, co-authors of this paper.

<sup>2</sup> The clinically significant change cutting points are those reported in Table 12 on page 58 of Evans *et al.* (2002), i.e. a criterion of  $\geq 1.36$  for men and  $\geq 1.50$  for women for the pro-rated mean of the 28 non-risk items and  $\geq 1.19$  for the men and  $\geq 1.29$  for the women on the pro-rated overall score. The reliable change criterion was based on the usual formula (see Evans, Barkham & Margison, 1998) using the baseline pro-rated non-risk standard deviation and coefficient alpha, this gave a criterion of a change of more than  $\geq 0.49$  for non-risk score for reliable change on the non-risk score and  $\geq 0.48$  for the overall score.

Table 1. Referential data for primary care counselling/therapy services

	Referential parameter value	Confidence interval (95%)
<i>Demographic parameters*</i>		
Gender		
Female	71.5%	70.4 to 72.6%
Ethnicity		
Non White-European	8.9%	8.2 to 9.7%
Age		
Under 20 years	4.5%	4.0 to 5.0%
Over 59 years	6.5%	5.9 to 7.1%
Employed (FT or PT)	55.6%	54.3 to 56.9%
Living alone	27.4%	26.3 to 28.6%
<i>Problem mix<sup>†</sup></i>		
Depression	73.3%	72.2 to 74.4%
Anxiety	79.3%	78.3 to 80.3%
Eating disorder	3.0%	2.6 to 3.5%
Addictions	5.6%	5.1 to 6.3%
Trauma/abuse	20.0%	19.0 to 21.1%
Interpersonal relationship problems	57.8%	56.5 to 59.0%
Living/welfare problems	15.0%	14.2 to 16.0%
Suicide risk <sup>‡</sup>	22.4%	21.4 to 23.5%
Self harm risk <sup>‡</sup>	21.4%	20.4 to 22.5%
Prescribed medication	48.5%	47.2 to 49.8%
<i>Intake severity (CORE-OM)<sup>§</sup></i>		
Overall CORE-OM item mean	1.81	1.79 to 1.83
Men	1.75	1.72 to 1.79
Women	1.83	1.81 to 1.85
Overall CORE-OM mean over clinical cut off <sup>¶</sup>	78.6%	76.9 to 80.1%
Men	77.5%	74.2 to 80.5%
Women	79.0%	77.0 to 80.8%
CORE-OM non-risk item mean	2.10	2.08 to 2.12
Men	2.02	1.98 to 2.06
Women	2.13	2.11 to 2.15
CORE-OM non-risk mean over clinical cut off <sup>¶</sup>	79.6%	78.0 to 81.1%
Men	79.7%	76.5 to 82.5%
Women	79.6%	77.7 to 81.4%
<i>Assessment outcome<sup>**</sup></i>		
Accepted for further sessions	88.4%	87.6 to 89.2%
Referred back to referrer or to other service	3.5%	3.1 to 4.0%
<i>Service parameters</i>		
Mean waiting time to first appointment (days)	46.4	45.0 to 47.9
Mean number of sessions offered	6.1	6.06 to 6.22
<i>Therapy outcome</i>		
Unplanned ending <sup>††</sup>	38.9%	37.5 to 40.4%
Final overall CORE-OM item mean	0.84	0.81 to 0.86
Men	0.87	0.82 to 0.92
Women	0.83	0.80 to 0.85
Final non-risk CORE-OM score	0.99	0.97 to 1.02
Men	1.03	0.97 to 1.08
Women	0.99	0.95 to 1.02
Initial and final CORE-OM both available <sup>††</sup>	38.0%	36.8 to 39.1%
Drop in overall item mean CORE-OM score	0.96	0.93 to 0.98
Men	0.89	0.83 to 0.94
Women	0.98	0.95 to 1.01
Rel. or clin. significant improvement	77.8%	76.1 to 79.4%
Men	73.8%	70.3 to 77.0%
Women	79.7%	77.4 to 81.1%
Clinically significant improvement	59.1%	57.1 to 61.0%
Men	54.7%	50.9 to 58.4%
Women	60.7%	58.4 to 62.9%

Table 1. Continued

Drop in non-risk CORE-OM score	1.10	1.07 to 1.13
Men	1.01	0.95 to 1.07
Women	1.13	1.09 to 1.16
Rel. or clin. significant improvement: non-risk	78.4%	76.7 to 79.9%
Men	74.6%	71.2 to 77.7%
Women	79.7%	77.8 to 81.5%
Clinically significant improvement: non-risk	59.3%	57.3 to 61.2%
Men	55.1%	51.4 to 58.8%
Women	60.8%	58.5 to 63.0%

\*Demographic variables *n* from 5589 to 6607.

† Practitioner-rated presenting problems from minimal to severe, total *n* = 5908.

‡ Practitioner-rated from mild to severe, total *n* = 5860 and 5825.

§ Total *N* = 5776 for non-risk scores and 5733 for all item scores.

¶ See Evans et al. (2002).

\*\* Total *n* = 6035 to 6325.

†† Total *n* = 4396.

‡‡ Total *n* = 2509 (based on non-risk item mean score).

appointment was just over 6 weeks and about 90% of clients were offered further sessions following assessment. About two in every five starting therapy terminated unilaterally. Outcome data showed four in five to have achieved reliable or clinically significant improvement.

These figures provide a summary profile for primary care-based counselling services using the CORE system and a benchmark against which services can review their own data. (The term ‘benchmark’ arises from the marks that tailors and carpenters made on their benches to make measurement quick and easy).

Services not using the data analysis and reporting services, but using the CORE measures can compare their service parameters to these benchmarks. This is shown in the next section.

## ROUTINE SERVICE FEEDBACK AND THE SERVICE/REFERENTIAL DATA COMPARISON

### Method

Each service sending their data for analysis was provided with a report comparing their data with that of an appropriate referential dataset. A summary of the report for one primary care counselling service (referred to as ‘service’), and which compares their data with the referential data outlined in the previous section, is presented here. The service is in the south of England serving a large multicultural population. Data collection, analyses and reporting for this service had occurred for 30 months with each new batch of data being added

to data previously received in order to maximize the amount of outcome data available.

The full report to the service comprised 34 pages and like that to all services using the PTRC batch processing system, contained simple tabular and graphical presentation of services’ data, descriptive statistics, means and percentages. Group summary presentations were supplemented with analyses of reliable and clinical change (Evans *et al.*, 1998; Jacobson & Truax, 1991) to provide analyses at the level of the individual client rather than the group or population. We believe that presenting both simple group summary and individual change analyses is more likely to engage practitioners’ interest and encourage further exploration of their data than are more complex statistical methods.

### Analyses

The analyses from the routine service report are summarized here and are supplemented with inferential tests and confidence intervals, and differences noted below are statistically significant at  $p < 0.05$  unless stated to the contrary. Any service can compare the confidence interval for its own parameters (see e.g. Gardner, Gardner, & Winter (1989) for a simple guide to calculation of confidence intervals from summary parameters). Where the referential value lies outside the confidence interval for the service, this is a fairly robust indicator of a statistically significant difference using the conventional criterion of 0.05. The reverse contrast, i.e. comparing a service parameter to the confidence interval for that parameter from the referential data in Table 1, is not a good way to test

whether the individual service's data differ significantly from that reference interval since the interval in question will be very tight given the large  $n$  on which it is calculated, i.e. for services with small  $n$  datasets the test would suggest a statistically significant difference far too often. This approach to comparison of service to referential data is illustrated in Table 2 below.

### Results

Table 2 shows the same variables as Table 1. For each variable it gives: the parameter for this particular service (column 2); the 95% confidence interval for the service parameter (column 3); the referential value (as in Table 1, column 4); and finally whether the referential lies outside that confidence interval for the service, i.e. whether the difference is statistically significant. The more notable differences and ones unlikely to have arisen through differences in therapist rating preferences, are shown in bold in column 5.

This shows that, compared to the rest of the referential data, the service sees more clients from ethnic minorities, a higher percentage in employment and a slightly higher percentage living alone and on medication and a slightly lower percentage of older clients. Clients attending the service had higher initial CORE-OM scores and more above the clinical cut-off score. The service offers therapy to slightly more of the referrals, but offering slightly fewer sessions (although it turned out that these are offered over a slightly longer period owing to not infrequently using fortnightly sessions). Notably, the service had a markedly shorter waiting time to starting therapy than the overall dataset value.

The results of greatest interest to the service concerned effectiveness (pre-therapy to post-therapy change). Analysis of reliable and clinical change showed that the service compared favourably overall with slightly, but non-significantly, higher rates of improvement by reliable and clinically significant change and large drops in scores overall and for women.

Services differ on many variables and each could potentially relate to outcomes and each warrants a more in-depth analysis than we are able to present in this paper. However, one area where the service differed most markedly from the referential dataset was in the proportion of clients from ethnic minorities and very little empirical survey data on ethnicity has been reported in the counselling

literature. Hence, the next section considers information on ethnicity in more detail.

### ETHNICITY AND ETHNIC MINORITIES

As noted above, there is a lack of research evidence relating to ethnic minority clients in the British primary care sector, even for basic parameters such as intake severity and outcomes of therapy. The data presented in this section illustrate how it is possible for a service to 'drill into' complexities of local interest such as ethnicity and how such analyses benefit from being placed in the context of national, referential data.

### Method

The data is derived from the datasets outlined in the previous two sections above. However, due to the relatively low number of sites returning ethnic minority data, the ethnic minority sample from the service dataset makes up a sizeable proportion of the referential data. Hence, in this section the service data is compared to the referential data, quoted in the previous two sections, *from which the service data has been removed*, called the 'comparison' dataset. Where appropriate, the values of the 'comparison' and 'service' datasets have been combined to give the reader an overall 'referential' value.

Ethnicity was categorized using the 'ethnic origin' section of the practitioner-completed CORE-A Therapy Assessment Form. The ethnicity categories used were consistent with the Department of Health classifications from the 1991 census (this was the most recent classification available when the work started). This gives categories of: Asian (Bangladeshi); Asian (Indian); Asian (Pakistani); Asian (East African); Asian (Chinese); Black (African); Black (Caribbean); White (English/European); Other. Two coding boxes were available allowing mixed or dual ethnicity to be noted. Most analyses below report on the contrast between 'White/European' (WE) and 'Ethnic Minority' clients (EM: defined as all not categorized as WE, including dual/mixed ethnicity). We recognize that grouping clients from different ethnicities has significant costs, hiding diversity of cultural backgrounds and religious beliefs in one group. However, the relatively small number of clients from some ethnic groups made pooling unavoidable for statistical analyses even in this size of dataset. Where statistically informative

Table 2. Comparison of this service with the referential data

	Service parameter	CI (95%) for service parameter	Referential parameter value (%)	Referential value outside service CI?
<i>Demographic parameters</i>				
Gender				
Female	68.8%	65.3 to 72.2%	71.5%	N
Ethnicity				
Non White-European	23.6%	20.5 to 27.0%	8.9%	Y
Age				
Under 20 years	5.5%	4.1 to 7.5%	4.5%	N
Over 59 years	4.2%	3.0 to 6.0%	6.5%	Y
Employed (FT or PT)	62.1%	58.3 to 65.7%	55.6%	Y
Living alone	29.4%	26.1 to 32.9%	27.4%	Y
<i>Problem mix</i>				
Depression	76.9%	73.6 to 79.9%	73.3%	Y
Anxiety	84.7%	81.8 to 87.2%	79.3%	Y
Eating disorder	2.0%	1.2 to 3.4%	3.0%	N
Addictions	5.7%	4.2 to 7.7%	5.6%	N
Trauma/abuse	28.1%	24.9 to 31.6%	20.0%	Y
Interpersonal relationship problems	68.7%	65.1 to 72.1%	57.8%	Y
Living/welfare problems	22.0%	19.0 to 25.2%	15.0%	Y
Suicide risk	21.8%	18.9 to 25.0%	22.4%	N
Self harm risk	23.3%	20.2 to 26.6%	21.4%	N
Prescribed medication	53.5%	49.7 to 57.2%	48.5%	Y
<i>Intake severity (CORE-OM)</i>				
Overall CORE-OM item mean				
Men	1.96	1.91 to 2.02	1.81	Y
Women	1.95	1.84 to 2.05	1.75	Y
Overall mean over clinical cut-off	1.97	1.90 to 2.03	1.83	Y
Men	85.7%	81.3 to 89.1%	78.6%	Y
Women	85.7%	77.4 to 91.2%	77.5	N
CORE-OM non-risk item mean	85.6%	80.2 to 89.8%	79.0%	Y
Men	2.25	2.19 to 2.31	2.10	Y
Women	2.22	2.11 to 2.33	2.02	Y
Non-risk mean over clinical cut-off	2.26	2.19 to 2.33	2.13	Y
Men	86.8%	82.6 to 90.1%	79.6%	Y
Women	85.9%	77.7 to 91.4%	79.7%	N
	87.3%	82.1 to 91.0%	79.6%	Y
<i>Assessment outcome</i>				
Accepted for further sessions	91.5%	89.1 to 93.3%	88.4%	Y
Referred back to referrer or to other service	2.3%	1.4 to 3.7%	3.5%	N
<i>Service parameters</i>				
Mean waiting time to first appointment (days)	27.4	24.8 to 30.0	46.4	Y
Mean number of sessions offered	5.6	5.5 to 5.7	6.1	Y
<i>Therapy outcome</i>				
Unplanned ending	41.5%	37.6 to 45.6%	38.9%	N
Final overall CORE-OM score				
Men	0.82	0.75 to 0.89	0.84	N
Women	0.89	0.74 to 1.03	0.87	N
Final non-risk CORE-OM score	0.79	0.71 to 0.88	0.83	N
Men	0.97	0.89 to 1.05	0.99	N
Women	1.04	0.88 to 1.20	1.03	N
Initial and final CORE-OM both available	0.94	0.85 to 1.04	0.99	N
Drop in overall CORE-OM score	45.0%	41.4 to 48.8%	38.0%	Y
Men	1.13	1.04 to 1.21	0.96	Y
Women	1.07	0.90 to 1.24	0.89	Y
Rel. or clin. significant improvement	1.16	1.06 to 1.26	0.98	Y
Men	81.1%	76.3 to 85.1%	77.8%	N
Women	76.5%	67.2 to 83.8%	73.8%	N
	83.3%	77.6 to 87.7%	79.7%	N



Table 2. *Continued*

Clinically significant improvement	64.2%	58.7 to 69.3%	59.1%	N
Men	59.2%	49.3 to 68.4%	54.7%	N
Women	66.5%	59.9 to 72.6%	60.7%	N
Drop in non-risk CORE-OM score	1.28	1.19 to 1.38	1.10	Y
Men	1.20	1.01 to 1.39	1.01	N
Women	1.32	1.21 to 1.44	1.13	Y
Rel. or clin. significant improvement	81.4%	76.6 to 85.3%	78.4%	N
Men	75.8%	66.5 to 83.1%	74.6%	N
Women	84.0%	78.4 to 88.3%	79.7%	N
Clinically significant improvement	64.3%	58.8 to 69.4%	59.3%	N
Men	58.6%	48.7 to 67.8%	55.1%	N
Women	67.0%	60.4 to 73.0%	60.8%	N

Table 3. Comparison of ethnic breakdown in service with comparison dataset

Primary ethnic group	Service	Comparison	Overall
Asian (Bangladeshi)	0 (0%)	27 (0.5%)	27 (0.5%)
Asian (Indian)	65 (9.8%)	53 (1.0%)	118 (2.0%)
Asian (Pakistani)	32 (4.8%)	29 (0.6%)	61 (1.1%)
Asian (East African)	30 (4.5%)	13 (0.3%)	43 (0.7%)
Asian (Chinese)	1 (0.2%)	10 (0.2%)	11 (0.2%)
Black (African)	6 (0.9%)	31 (0.6%)	37 (0.6%)
Black Caribbean	11 (1.7%)	83 (1.6%)	94 (1.6%)
White (English European)	505 (76.4%)	4741 (93.0%)	5246 (91.1%)
Other (incl. dual ethnicity)	11 (1.7%)	100 (2.1%)	121 (2.1%)
Total with recorded ethnicity	661 (100.0%)	5097 (100.0%)	5758 (100.0%)

First entry in each row is *n* for cell, followed by the percentage.

comparisons could be made between smaller ethnic minority groups formed by pooling: e.g. Asian Bangladeshi with Pakistani; Asian Indian with East African; Black African with Caribbean, these are reported. The Asian Chinese, Mixed race and 'Other' categories were not included in these subgroup analyses.

## Results

### *Ethnicity Breakdown*

A first ethnicity group was given for 661 of the 690 clients seen in the service (95.8%), a much higher percentage than for the comparison dataset (5099 of 5920, 86.1%,  $p < 0.0005$ ). The breakdown of first ethnic group is shown in Table 3. This service reported more clients in the Asian Indian, Pakistani and East African groups with a corresponding decrease in the number in the 'White/European' group. For some groups, the '*n*' in the service data is larger than that in the much larger comparison dataset. In itself, this illustrates how some services may have specific issues to explore.

### *Demographic and Presenting Problem Differences*

The gender balance did not differ statistically significantly between the EM and WE samples either for this service (69.9 vs. 69.7% female respectively) or within the referential datasets (71.1 vs. 72.2%). The smaller groups also revealed no major differences with the percentage female ranging within groups within the service from 63% ( $n = 20$ , Asian Pakistani/Bangladeshi) to 73% ( $n = 69$ , Indian/East African). Similarly, there were no statistically significant differences between ethnic groups in age by ethnic group (mean from 33 to 37 years) within the service but there were statistically significant differences in the comparison dataset (K-W test  $p = 0.002$ ) with Pakistani/Bangladeshi mean age 33 years, Indian/East African mean age 36 years, Black African/Caribbean 38 years and White/European 39 years, differences that may reflect both population age differences and ethnically-specific differences in willingness to seek or accept psychological therapies with age, or differences in willingness of GPs to refer by ethnicity.

Table 4. Comparison of ethnic minority and white European on pre-therapy CORE outcome dimension mean scores for service and comparison data

	Ethnic minority		White European		Difference		
	Mean	SD	Mean	SD	<i>P</i>	Effect size	95% CI diff.
<b>Comparison services*</b>							
Well-being	2.36	0.90	2.37	0.87	0.979	-0.01	-0.02 to 0
Symptoms/problems	2.36	0.86	2.28	0.80	0.094	0.1	0.09 to 0.11
Functioning	1.93	0.79	1.77	0.76	0.001	0.21	0.20 to 0.22
Risk	0.49	0.58	0.44	0.62	0.255	0.08	0.07 to 0.09
All items excluding risk	2.18	0.76	2.07	0.72	0.017	0.15	0.15 to 0.16
Total mean score	1.88	0.68	1.78	0.66	0.019	0.15	0.14 to 0.16
<b>Service†</b>							
Well-being	2.54	0.93	2.49	0.86	0.589	0.06	0.04 to 0.08
Symptoms/problems	2.63	0.91	2.39	0.80	0.003	0.29	0.27 to 0.31
Functioning	2.18	0.94	1.90	0.73	0.002	0.36	0.34 to 0.38
Risk	0.74	0.71	0.57	0.70	0.015	0.24	0.22 to 0.26
All items excluding risk	2.44	0.87	2.20	0.71	0.005	0.32	0.30 to 0.34
Total mean score	2.14	0.81	1.91	0.66	0.003	0.33	0.31 to 0.35

\*Ethnic minority *n* from 279 to 291, white European *n* from 3953 to 4066.

†Ethnic minority *n* from 124 to 130, white European *n* from 464 to 480.

Ethnic minority clients in this service were less likely to be employed than the local WE clients (56 vs. 64%, *p* = 0.07) a slightly smaller difference than in the comparison data (49 vs. 55%, *p* = 0.03). Service EM clients were less likely to be living alone (21 vs. 33%, *p* = 0.004), but for the comparison dataset, EM clients were slightly more likely to be living alone (27.5 vs. 27%, *p* = 0.85). The EM and WE clients showed considerable and highly statistically significant differences in proportions with the various identified problems covered in the CORE-A TAF forms, specifically they were significantly more likely to be noted to have problems of: bereavement/loss (61 vs. 30%, *p* < 0.0005), depression (89 vs. 74%, *p* < 0.0005), interpersonal relations (78 vs. 66%, *p* = 0.003), work/academic nature (32 vs. 19%, *p* = 0.001), personality (30 vs. 12%, *p* < 0.0005), physical (55 vs. 23%, *p* < 0.0005), trauma/abuse (49 vs. 22%, *p* < 0.0005), self-esteem (79 vs. 57%, *p* < 0.0005) and anxiety (90 vs. 83%, *p* = 0.04); but there were no differences between the two groups on problems of addictions, living/welfare, cognitive/learning disability, eating disorders (very low rates, 1–2%, in both groups), psychoses (ditto).

The practice expressed interest in some particular questions as described below.

*Practice-based Question: Are There Differences in Intake Severity Between Ethnic Minority and White European Clients?* Ethnic minority clients in the comparison data scored significantly differently

(higher, i.e. more distressed) only on the functioning CORE-OM domain score. Within the service data, EM clients scored significantly higher than WE on all scores except well-being (see Table 4).

*Practice-based Question: Are There Differences in Severity Within the Ethnic Minority Group at Intake?*

Table 5 shows a comparison of pre-therapy CORE-OM mean scores by ethnicity for service and comparison data. The differences within groups in the table are statistically significant (K-W test *p* = 0.014) but non-significant if the WE group is removed (*p* = 0.11). The overall referential data, i.e. the service and comparison data combined, are provided in the final columns for services wishing to contextualize their own data. The data show somewhat significantly higher pre-therapy mean scores for the service than comparison services for the WE group and a difference in the opposite direction that is nearly statistically significant for the Asian Indian/East African group. This difference is shown more clearly in the notched boxplot of the same data in Figure 1 on which the overall mean score is plotted as the same reference line for each subplot. It can be seen that in this service the scores for the WE and Black African/Caribbean groups both seem to have 95% confidence intervals (notches on the boxes) not including the overall referential mean and that the Indian/East African group have a markedly higher median score.

Table 5. Comparison of pre-therapy CORE-OM non-risk mean scores by ethnicity for primary care counselling service and comparison data

Ethnic group	Service			Comparison			95% CI difference	Overall referential		
	N	Mean	95% CI	N	Mean	95% CI		N	Mean	95% CI
Asian Pakistani/ Bangladeshi	22	2.06	1.72 to 2.4	43	2.23	1.99 to 2.47	-0.58 to 0.24	65	2.17	1.97 to 2.37
Asian Indian/ East African	83	2.51	2.31 to 2.71	58	2.14	1.96 to 2.32	0.09 to 0.65	140	2.36	2.22 to 2.50
Black African/ Caribbean	15	2.40	2.03 to 2.77	92	2.11	1.93 to 2.29	-0.18 to 0.76	106	2.14	1.99 to 2.31
White/European	482	2.20	2.14 to 2.26	4066	2.07	2.05 to 2.09	0.06 to 0.20	4503	2.09	2.07 to 2.11

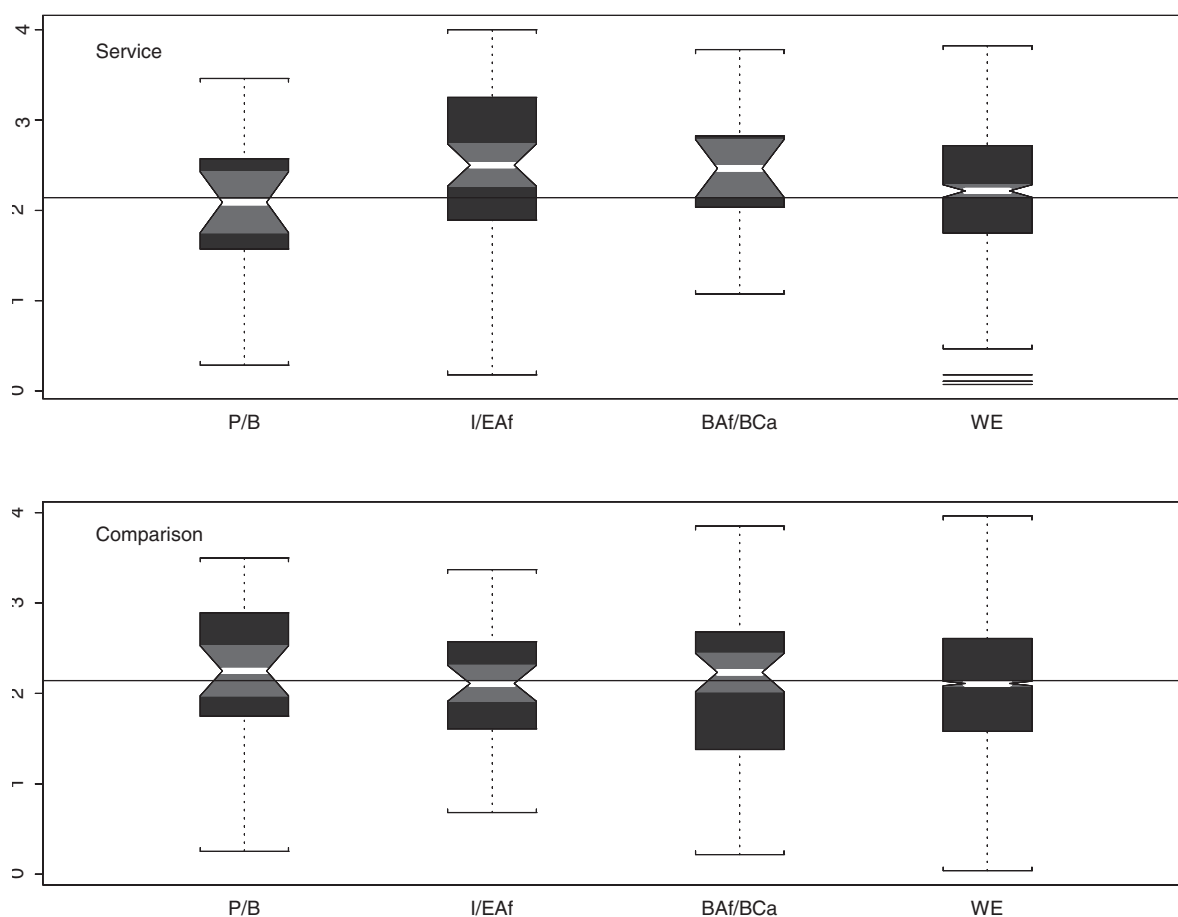


Figure 1. Boxplot of CORE-OM scores by service/comparison and ethnic subgroups. P/B, Pakistani/Bangladeshi; I/Eaf, Indian/East African; BAf/BCa, Black African/Black Caribbean; WE, White European. Horizontal reference line is the overall median of all groups' data combined. Typical boxplot conventions have been followed. Boxes contain the inter-quartile interval of the score distribution, whiskers extend to maxima and minima although outliers beyond 1.5× inter-quartile distance are shown as individual horizontal lines beyond the whiskers. Notches give 95% confidence interval of the median (the waste of each box)

*Practice-based Question: Are Clients from Ethnic Minorities more Likely to Drop out of Therapy?* Analysis was carried out for those clients who had an unplanned ending to therapy as recorded on the End of Therapy Form. Ethnic minority clients in the service were more likely to have an unplanned ending than 'White/European' clients, 47 and 39% respectively, although this was not statistically significant ( $p = 0.15$ ). There was scarcely any difference between the two groups in the comparison data (38.4 vs. 38.1%). However, this analysis conceals a more complex picture when analysed by the ethnic groupings within the service when a statistically significant difference ( $p = 0.02$ ) emerges with high early termination in the Pakistani/Bangladeshi (66%) and Black African/Caribbean (59%) groups cf. 40% for Indian/E.African and 39% for WE. The differences in the larger comparison dataset were very small and not statistically significant.

*Practice-based Question: Are There any Differences in Therapy Outcome Between EM and WE Clients?* Analysis of pre- to post-therapy change on the CORE-OM by ethnicity showed that for the service there was no association between ethnicity and improvement (ANCOVA of final non-risk CORE-OM score with assessment score as covariate:  $F(1;301) = 0.30, p = 0.59$ ) whereas a small difference for the comparison services was narrowly statistically significant (ANCOVA  $F(1;1878) = 4.1, p = 0.04$ ). Entering the entire referential dataset into a two-way ANCOVA against ethnicity and service (this service cf. comparison) showed significant simple effects for service ( $F(1;2180) = 7.4, p = 0.007$ ) but not ethnicity ( $F(1;2180) = 2.1, p = 0.15$ ), but the interaction, although showing a much smaller advantage to the WE group in the WE/EM contrast in the service, was not statistically significant ( $F(1;218) = 0.56, p = 0.46$ ).

## DISCUSSION

This paper illustrates the first close collaboration between the CORE system data analytic service and a counselling practice. As well as presenting substantive findings we have sought to show how the PBE paradigm can complement EBP and create a knowledge base useful at the national level, within services, and, in principle, even at the level of the individual practitioner.

The first key issue is that the growth of the referential dataset puts things in context and enables services to see ways in which they differ from other services. This is shown in the first section where we

report a range of referential service data across primary care settings. These reflect benchmarks on these parameters from a large number of primary care-based psychological therapy services and their clients. These are benchmarks, not standards set by government or other organization, and they have not been generated in response to political or managerial target setting. They simply reflect practice as it is in the range of routine clinical services participating in the data collection. The dataset is wholly different from that utilized by Barkham *et al.* (2001) which focused on secondary and specialist services, and it considerably enlarges that reported in Mellor-Clark, Connell, Barkham, and Cummins (2001) though it contains 2802 client records from that report. A number of findings are very similar across these three surveys suggesting the robustness of certain parameters to changes of setting and cohort. For example, the percentage of people scoring above the clinical threshold at intake is fairly consistent around 80%, a finding that suggests that the remaining 20% may be worthy of specific study.

In the second section, we looked at the analysis provided for one service using the CORE data reporting service, comparing their data with the referential data reported in the first section. Table 2 showed how a service could look at its own parameters and their confidence interval. If the referential value lies outside the confidence interval of the service's own data, a formal test comparing that service with the referential data would almost certainly have shown a statistically significant difference. Hence a similar process of statistical comparison is possible for any service willing to collect their own data and do some calculations.

Services using CORE-PC have the parameters computed for them and can easily add confidence intervals to that. Comparison with the referential data showed this particular service that it took less than the national average time from assessment to first session; that they offered slightly less than the national norm for numbers of sessions, but spread those over longer therapy duration through the frequent use of fortnightly rather than weekly therapy. The referential data allowed the service to see that the outcomes for their clients were similar to, or slightly better than, other services in CORE-OM score improvement. Furthermore, simple analyses suggested that the service was seeing clients with equivalent or slightly greater problem severity than other services.

The third section illustrates how any service with a reasonable  $n$  in its own accumulating dataset can

focus on issues of local interest; how the PBE paradigm allows exploration of issues requiring dataset sizes that are currently not available from more traditional EBP data collection. For this service, the topic of most interest was client ethnicity. Results showed rather few differences between ethnic groups although there seemed to be significant differences between ethnic groups in rates of early termination and very marked differences in the therapists' ratings of presenting problems. There was a very small difference between EM and WE clients in improvement on the CORE-OM where both assessment and end of therapy scores were available.

Clearly some caution is needed in interpreting differences on CORE-OM scores. Although the CORE-OM showed no internal consistency differences between British university students whose first language was not English and those for whom it was (Evans *et al.*, 2002), we do not yet have extensive normative data for different ethnic groups outside university settings. Another reason for caution is that ethnicity is highly complex and effects can be specific to ethnicity differences lost when smaller ethnic groups are pooled to provide a WE/EM contrast with sufficient *n* in the smaller group to meet statistical power concerns. This loss of smaller group effects by pooling is shown here for early termination of therapy. Yet another complication is that ethnicity (and other) effects can be service specific and it is notable that this service employs counsellors of Asian Indian origin. As their dataset accumulates, further investigation could be undertaken by this service, of effects of counsellor ethnicity and of ethnic matching of client and counsellor.

Although we are highlighting some evidence that PBE can give that EBP methods cannot, our main argument remains that PBE and EBP are both necessary and complementary (see Margison *et al.*, 2000). Thus PBE does not answer issues of causal attribution explored in EBP RCTs but complements findings from such studies by providing evidence of comparability of typical RCT change scores to those seen in routine practice. In addition, PBE provides a quasi-epidemiological evidence base about the demographics of referred populations on variables such as client ethnicity, that cannot be randomized, and other things, such as counsellor ethnicity, that will never or rarely be experimentally controlled for logistic or ethical reasons. Such PBE is only possible through the creation of large referential datasets and local accumulation of service data.

Previously, it has been rare for psychological therapy services to collect moderately extensive standard data on all clients and rarer still for such data to be other than managerial information about progress through 'the system'. The CORE system provides self-report measures that seem acceptable to clients/patients and of good psychometric properties and a useful therapist-completed measure for assessment and end of therapy. However, none of the principles and procedures exemplified here are measure specific. One unique feature of the CORE system at present however, is that the psychometrically good measures are copyleft (i.e. copyright but free to reproduce providing no changes or profits are made). Free and readily available measures and proformas for therapist completion are one step forward. However, without intelligent analysis of the data there is the danger of perpetuating the 'bottom drawer' phenomenon whereby services collect data which is not utilized.

To minimize this risk, the data collected were designed to be minimal, a 'core', onto which other measures, appropriate to specific problems or to local services, would be added. However, even this 'core' dataset is complex, with many client variables and service parameters likely to be correlated strongly with others. We have been working to find ways of presenting the batch-processed data to services in ways that they would find useful. Likewise, services have been working to find the best ways to take on board the results and combine them with detailed qualitative information about clients from the clinical sessions and with their own knowledge of local demography, services and service politics. For PBE to have a real impact on practice, such datasets need to be analysed and described and clinicians and service managers to feel comfortable with such analyses and to be comfortable with contextualizing local analyses against referential data. Once this is achieved, PBE can take service quality improvement beyond audit, toward evaluation and identification of best practice and how that is achieved, something that will require local questions and exploration, not just national prescriptions and our choice, clear to the services purchasing the batch processing, is for comparison against the rest of the referential data, not named or anonymized ranking or 'star chart' systems.

There are caveats. Although the largest referential data in existence on primary care psychological therapies in the UK to our knowledge, such data are, necessarily, influenced by the biases that

lead some services to participate in such data collection and others not to do so. Service participation undoubtedly results in a non-random selection of all primary care psychological therapy services in the country although the selection bias may be less than that determining participation in RCTs in the psychological therapies.

Even within the services participating, there is more selectivity, and outcome data is available for fewer than half the clients. We are working on the viability of statistical methods of complementing self-report data with the therapist ratings of change on the TAF and EOT forms of the CORE-A to provide information where the self-report data are missing. However, such methods are only likely to succeed if there are large-scale inter-rater reliability training which will be very expensive. Some services currently use frequent repetition of shorter self-report forms (CORE-Short Forms A and B) so that a trajectory to termination is visible where clients terminate early (Barkham *et al.*, 2001). Others use the CORE-OM at monthly or 3-monthly intervals in longer therapies for the same purpose, and these patterns of data collection currently provide the best partial solution to the loss of data from termination pending record linkage or routine postal or telephone follow-up enabling review of mental state and service use at intervals after termination.

Whether these various biases make PBE datasets more or less prone to generalizability problems than are efficacy RCTs, or even pragmatic RCTs, is unclear and something that can only be answered by new research looking carefully at these effects and at the effects of study exclusion criteria. We hope, within the next 2 years, to report data from an entirely distinct primary care dataset of services using the CORE-PC direct data entry system which may throw light on at least some selection effects.

However, the biggest challenge to real quality improvement in psychological therapies probably remains that of bridging practitioner–researcher gaps, hence researchers and data analysts need to find the most appropriate ways to present data and analyses; and equally, practitioners need to become increasingly confident and influential in guiding the analyses of data. We hope this paper has given some indication of possible ways forward in these respects.

## ACKNOWLEDGEMENTS

This paper would not have been possible without the collaboration of members of the service who

have chosen to remain anonymous, equally the referential dataset would not have come into existence without the collaboration of hundreds of professionals across the UK. Early work was funded by the Artemis Trust and the Mental Health Foundation. JC, MB CM and JM-C acknowledge support from Leeds Community and Mental Health Trust and CE support from Nottinghamshire Healthcare NHS Trust and the Tavistock & Portman NHS Trust.

## REFERENCES

- Barkham, M., & Mellor-Clark, J. (2000). Rigour and relevance: practice-based evidence in the psychological therapies. In N. Rowland, & S. Goss (Eds), *Evidence-based health care in psychological therapies* (pp. 127–144). London: Routledge.
- Barkham, M., Evans, C., Margison, F., McGrath, G., Mellor-Clark, J., Milne, D., & Connell, J. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health*, 7, 35–47.
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, 69, 184–196.
- Bentovim, A., Boston, P., & Van Elburg, A. (1987). Child sexual abuse—children and families referred to a treatment project and the effects of intervention. *British Medical Journal*, 295, 1453–1457.
- Bower, P., Richards, D., & Lovell, K. (2001). The clinical and cost-effectiveness of self-help treatments for anxiety and depressive disorders in primary care: a systematic review. *British Journal of General Practice*, 51, 838–845.
- Brooker, C., Falloon, I., Butterworth, A., Goldberg, D., Graham-Hole, V., & Hillier, V. (1994). The outcome of training community psychiatric nurses to deliver psychosocial intervention. *British Journal of Psychiatry*, 165, 222–230.
- Colyer, H., & Kamath, P. (1999). Evidence-based practice—A philosophical and political analysis: some matters for consideration by professional practitioners. *Journal of Advanced Nursing*, 29, 188–193.
- CORE System Group. (1999). *CORE system user manual*. Leeds, UK: Psychological Therapies Research Centre, University of Leeds.
- Daryanani, R., Hindley, P., Evans, C., Fahy, P., & Turk, J. (2001). Ethnicity and use of a child and adolescent mental health service I: ethnic breakdown of referrals. *Child Psychology and Psychiatry Review*, 6, 127–132.
- Department of Health. (1996). *NHS psychotherapy services in England: review of strategic policy*. London: HMSO.
- Department of Health. (2001). *Treatment choice in psychological therapies and counselling*. London: HMSO.

- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*, 51–60.
- Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health*, *1*, 70–72.
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health*, *9*, 247–255.
- Eynon, T. (2001). Narcissism and dyslexia: a patient with specific learning difficulties in time limited psychotherapy. *Journal of Critical Psychology, Counselling and Psychotherapy*, *1*, 94–102.
- Faulkner, A., & Thomas, P. (2002). User-led research and evidence-based medicine. *British Journal of Psychiatry*, *180*, 1–3.
- Ford, M.F., Jones, M., Scannell, T., Powell, A., Coombes, R.C., & Evans, C. (1990). Is group psychotherapy feasible for oncology outpatients attenders selected on the basis of psychological morbidity? *British Journal of Cancer*, *62*, 624–626.
- Freud, S. (1977). *Case histories I. 'Dora' and 'Little Hans'*. Harmondsworth: Penguin Books Ltd.
- Gardner, M.J., & Altman, D.G. (1986). Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal*, *292*, 746–750.
- Gardner, M.J., Gardner, S.B., & Winter, P.D. (1989). *Confidence Interval Analysis (C.I.A.) microcomputer program manual*. London: British Medical Journal.
- Gilbody, S.M., House, A.O., & Sheldon, T.A. (2001). Routinely administered questionnaires for depression and anxiety: systematic review. *British Medical Journal*, *322*, 406–409.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.
- Kohut, H. (1979). The two analyses of Mr. Z. *International Journal of Psycho-Analysis*, *60*, 3–27.
- Kramers, M., Evans, C., Dolan, B., Hume, F., & Lacey, H. (1991). The changing face of bulimia. *British Review of Bulimia and Anorexia Nervosa*, *5*, 17–21.
- Lambert, M.J., & Ogles, B.M. (2003). The efficacy and effectiveness of psychotherapy. In M.J. Lambert (Ed.), *Garfield and Bergin's handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley (in press).
- Lambert, M.J., Whipple, J.L., Smart, D.W., Vermeersch, D.A., Nielsen, S.L., & Hawkins, E.J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: are outcomes enhanced? *Psychotherapy Research*, *11*, 49–68.
- Luborsky, L., McLellan, A.T., Woody, G.E., O'Brien, C.P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, *42*, 602–611.
- Margison, F.R., Barkham, M., Evans, C., McGrath, G., Mellor-Clark, J., Audin, K., & Connell, J. (2000). Measurement and psychotherapy: evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, *177*, 123–130.
- Marshall, M. (2002). Randomised controlled trials—misunderstanding, fraud and spin. In S. Priebe, & M. Slade (Eds), *Evidence in mental health care* (pp. 59–71). Hove: Brunner-Routledge.
- McCabe, R., & Priebe, S. (2002). Focusing on quality of life in treatment. *International Review of Psychiatry*, *14*, 225–230.
- Mellor-Clark, J., Barkham, M., Connell, J., & Evans, C. (1999). Practice-based evidence and need for a standardised evaluation system: informing the design of the CORE System. *European Journal of Psychotherapy, Counselling and Health*, *3*, 357–374.
- Mellor-Clark, J., Connell, J., Barkham, M., & Cummins, P. (2001). Counselling outcomes in primary health care: a CORE system data profile. *European Journal of Psychotherapy, Counselling and Health*, *4*, 65–86.
- National Advisory Mental Health Council, National Institute for Mental Health. (1999). *Bridging science and service: a report by the National Advisory Mental Health Council's Clinical Treatment and Services Research Workshop*. (NIH publication no. 99-4353). Washington DC: NIH.
- Parry, G.D., Shapiro, D.A., & Firth, J. (1986). The case of the anxious executive: a study from the research clinic. *British Journal of Medical Psychology*, *59*, 221–233.
- Roth, A., & Fonagy, P. (1996). *What works for whom?: A critical review of psychotherapy research*. New York: Guilford.
- Rothman, K.J. (1978). A show of confidence. *New England Journal of Medicine*, *299*, 1362–1363.
- Rowland, N., Godfrey, C., Bower, P., Mellor-Clark, J., Heywood, P., & Hardy, R. (2000). Counselling in primary care: A systematic review of the research evidence. *British Journal of Guidance & Counselling*, *28*, 215–231.
- Ryle, A. (1980). Some measures of goal attainment in focused integrated active psychotherapy: a study of fifteen cases. *British Journal of Psychiatry*, *137*, 475–486.
- Treasure, J., Schmidt, U., Troop, N., Tiller, J., Todd, G., Keilen, M., & Dodge, E. (1994). First step in managing bulimia nervosa: controlled trial of therapeutic manual. *British Medical Journal*, *308*, 686–689.
- Trinder, L. (2000). A critical appraisal of evidence-based practice. In L. Trinder with S. Reynolds (Eds), *Evidence-based practice: A critical appraisal*. Oxford: Blackwell Science Ltd.
- Tustin, F. (1958). Anorexia nervosa in an adolescent girl. *British Journal of Medical Psychology*, *31*, 184–200.
- Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., Gabbay, M., Tarrrier, N., & Addington-Hall, J. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *British Medical Journal*, *321*, 1383–1388.